JOURNAL OF COMPLEXITY 6, 199–230 (1990)

On Average Case Complexity of Linear Problems with Noisy Information*

L. PLASKOTA

Institute of Informatics, University of Warsaw, 00-901 Warsaw, Poland

Received April 3, 1989

We present general results on the average case complexity of approximating linear operators when only noisy information is available. We assume that the space of a linear problem is a Banach space equipped with a Gaussian measure μ. The error of evaluating linear functionals is assumed to be a Gaussian random variable with mean zero and variance depending on the measure μ. Any linear functionals and repetitive evaluations of them are permitted. The formulas for the optimal algorithm, nth optimal radius, and information, as well as tight bounds on the average ε-complexity, are given. A major result is that the effect of noisy information on average ε-complexity is completely characterized by whether or not the average ε-complexity for exact information goes to infinity faster than ε^{-2}. In particular, solving *any* nontrivial problem with noisy information has complexity proportional to at least ε^{-2}. © 1990 Academic Press, Inc.

1. INTRODUCTION

The average case complexity of linear problems is the minimal average cost of computing approximate solutions with average error no greater than ε. A substantial portion of the monograph of Traub *et al.* (1988), "Information-Based Complexity," is devoted to this subject and contains references to many papers. We will refer to this monograph as IBC. In most of these papers it is assumed that information about problem elements is computed exactly. Far fewer papers deal with noisy information from a complexity viewpoint. One of two assumptions is made. The first assumes the worst case setting with uniformly bounded noise. This approach is presented in, e.g., Kacewicz *et al.* (1986), Lee *et al.* (1987), Marchuk and Osipenko (1975), Melkman and Micchelli (1979), Micchelli and Rivlin (1977), Milanese and Tempo (1984), Viccino *et al.* (1987), and Wasilkowski (1987).

* Invited Paper.

0885-064X/90 $3.00

The alternative assumption is the average case setting with stochastic noise. Kadane *et al.* (1989) study infinite dimensional linear problems. Linear problems have also been extensively studied in Bayesian statistics. As a rule, the noise is assumed to be "white"; i.e., the error of evaluating a functional L is a Gaussian random variable with mean zero and variance independent of L. It is known, under some assumptions on the prior distribution, that smoothing splines are the best estimators of stochastic processes; see, e.g., Kimeldorf and Wahba (1970) and Wahba (1984). Many papers are devoted to the optimal choice of the smoothing parameter which controls the trade-off between infidelity to the data and roughness of the estimated function. These papers include Craven and Wahba (1979), Ragozin (1983), Reinsch (1971), Speckman (1985), Wahba (1975, 1985), and Wahba and Wang (1987). In a Bayesian framework, optimal information for finite dimensional problems is analyzed in Chaloner (1984), where the geometric interpretation of optimal designs is given. Optimal designs in regression models are studied in, e.g., Fedorov (1972), Kiefer (1959), Kiefer and Wolfowitz (1959), O'Hagan (1978), Sacks and Ylvisaker (1966), and Wahba (1971). A relation between Bayesian statistics and information-based complexity is discussed in Kadane and Wasilkowski (1985).

We think that the assumption that the variance is independent of the functional and measure is not always appropriate, at least for infinite dimensional spaces. To explain this, consider, for example, the classical Wiener measure on the space of continuous functions $f\colon [0, 1] \to \mathbb{R}$, where $f(0) = 0$. Since f is continuous and $f(0) = 0$ it seems quite natural to demand that one can observe the value $L(f) = f(x)$, $0 \le x \le 1$, more precisely for x close to zero than for x close to one. This is not compatible with the assumption that the variance of the noise is constant.

Therefore in this paper we analyze another kind of noise. Namely, we assume that the variance of the noise is equal to $\delta\|L\|_\mu^2$, where the μ-norm of L is given by $\|L\|_\mu = \sqrt{L(C_\mu L)}$. Here C_μ is the correlation operator of the measure μ placed on Borel sets of the space F of a linear problem, $C_\mu\colon F^* \to F$. For the classical Wiener measure we have $L(C_\mu L) = x$, for $L(f) = f(x)$, and the variance is equal to δx. Thus, the variance is smaller for x closer to zero. In particular, for $x = 0$ the value $f(x)$ is known exactly.

The assumption that the variance of the noise is $\delta\|L\|_\mu^2$ allows us to obtain results more specific than those obtained by Kadane *et al.* (1989), who present a more general approach. Note that in finite dimensional spaces the assumption that the variance is equal to $\delta\|L\|_\mu^2$ corresponds to the assumption that the variance is equal to $\delta\|L\|^2$, if C_μ is taken as the identity operator.

We believe that the variance of the noise should depend not only on L

and μ but also on the exact value $L(f)$. This type of noise will be the topic of future research.

We now outline the contents of this paper. In this paper by *a linear problem in the average case setting with noisy information* we mean the following: Let $S: F \to G$ be a continuous linear operator from a separable Banach space F to a separable Hilbert space G. The space F is equipped with a Gaussian measure μ. We want to approximate Sf based on noisy evaluations of $L_1(f), L_2(f), \ldots$, where $L_i \in F^*$ are continuous linear functionals. We stress that L_i may be adaptively chosen depending on the previously computed information. The noisy evaluation of $L_i(f)$ means that we have $z_i = L_i(f) + x_i$, where x_i is a Gaussian random variable with mean zero and variance equal to $\delta\|L_i\|_\mu^2$, $\|L_i\|_\mu = \sqrt{L_i(C_\mu L_i)}$, where $\delta \geq 0$ and C_μ is a positive definite correlation operator of the Gaussian measure μ, $C_\mu: F^* \to F$.

We first show how to combine noisy information ($z = [z_1, z_2, \ldots, z_n]$ about f) to get approximations to Sf with minimal average error (which is called the average radius of N). Then we study the optimal nonadaptive choice of n functionals L_i which form information N. That is, we construct the nth optimal information and find its radius (which is called the nth optimal radius). The formula for the nth optimal radius is given in terms of δ and the eigenvalues of the correlation operator C_ν of the a priori measure $\nu = \mu S^{-1}$ on the space G of solution elements. In the presence of noise, $\delta > 0$, the nth optimal radius tends to zero no faster than $n^{-1/2}$.

Then we consider adaptive information. In this case the results of Wasilkowski (1986) for exact information may be directly carried over to the noisy case. From this we conclude that adaption is not more powerful than nonadaption for linear problems with noisy information.

Tight lower and upper bounds on the average ε-complexity are obtained. These bounds enable us to infer the effect of noise on ε-complexity. This effect is completely characterized by the ε-complexity for exact information. The noise does not affect the ε-complexity whenever the ε-complexity for exact information, i.e., for $\delta = 0$, tends to infinity faster than ε^{-2} as $\varepsilon \to 0$. On the other hand, if the ε-complexity for exact information tends to infinity slower than ε^{-2} then the existence of the noise, $\delta > 0$, increases the ε-complexity to be proportional to $\delta\varepsilon^{-2}$. This holds for arbitrary $S \neq 0$.

We conclude our paper by an example. We show that for an approximation problem the least-squares algorithm, although not optimal, produces approximate solutions with almost minimal cost.

Some of the results concerning, for instance, adaptive information (Theorem 6.1) or the ε-complexity (Theorem 7.1 and Corollary 7.1) are true also in the more general case when the set Λ of permissible functionals is a proper subset of F^*. Unfortunately, it is not easy to find the nth optimal radius and information in this general case, even for S being a functional. They are very sensitive to the structure of Λ. We hope, however, that it is possible to successfully study noisy information with a restricted set of permissible functionals for such linear problems as integration or function approximation.

2. Basic Definitions

Let F be a real separable Banach space equipped with a Gaussian measure μ defined on Borel sets. We assume that μ has mean zero and a positive definite correlation operator C_μ: $F^* \to F$. The measure μ reflects our belief as to how often subsets of F occur. We wish to approximate the continuous linear solution operator

$$S: F \to G,$$

where G is a real separable Hilbert space with an inner product $\langle \circ, \circ \rangle$ and corresponding norm $\|\circ\|$. We assume that an element f, for which we want to approximate Sf, is not known exactly. We can, however, observe noisy information on f. More precisely, consider a (nonadaptive) information operator N: $F \to \mathbb{R}^n$ of the form

$$N(f) = [L_1(f), L_2(f), \ldots, L_n(f)], \qquad \forall f \in F,$$

where L_i are nonzero continuous linear functionals, $L_i \in F^*$, $i = 1, 2, \ldots, n$. Noisy information means that the evaluations of $L_i(f)$ are erroneously computed (observed). That is, instead of $N(f)$ we observe a vector $z = [z_1, z_2, \ldots, z_n]$, where $z_i = L_i(f) + x_i$, $i = 1, 2, \ldots, n$. The vector $x = [x_1, \ldots, x_n]$ denotes the noise. Our assumption about the noise is that x_i are independent Gaussian random variables with mean zero and variance σ_i^2 depending only on a parameter δ and the μ-norm of L_i, i.e.,

$$\sigma_i^2 = \delta \cdot \|L_i\|_\mu^2, \qquad i = 1, 2, \ldots, n,$$

where $\|L\|_\mu = \sqrt{L(C_\mu L)}$, $\forall L \in F^*$.

Hence, given N and $f \in F$ the probability that the observed z is in a Borel set B is

$$\operatorname{Prob}(z \in B) = \pi_\delta(B \mid f, N), \qquad \forall f \in F,$$

where the measure π_δ is given by

$$\pi_\delta(B \mid f, N) = (2\pi)^{-n/2}(\sigma_1 \ . \ . \ . \ \sigma_n)^{-1} \int_B \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{(t_i - L_i(f))^2}{\sigma_i^2}\right\}$$

$$dt_1 dt_2 \ . \ . \ . \ dt_n, \qquad \text{for } \delta > 0,$$

and

$$\pi_0(B \mid f, n) = \begin{cases} 1 & \text{if } N(f) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

An approximation to S is provided by an algorithm ϕ that uses N. More precisely, ϕ is a transformation

$$\phi\colon \mathbb{R}^n \to G.$$

The goodness of ϕ is determined by its average error, defined as

$$e_\delta(\phi, N) = \left\{\int_F \left(\int_{\mathbb{R}^n} \|Sf - \phi(z)\|^2 \pi_\delta(dz \mid f, N)\right) \mu(df)\right\}^{1/2}.$$

We stress that the error $e_\delta(\phi, N)$ depends not only on N and ϕ, but also on the parameter δ. For $\delta = 0$ we observe the value $N(f)$ with probability one for every $f \in F$, and the error of ϕ takes the form

$$e_0(\phi, N) = \left\{\int_F \|Sf - \phi(N(f))\|^2 \mu(df)\right\}^{1/2}.$$

We are interested in information operators N and algorithms ϕ for which the error $e_\delta(\phi, N)$ is as small as possible.

3. Radius of Information and Optimal Algorithm

For an information operator $N\colon F \to \mathbb{R}^n$, $N = [L_1, L_2, \ . \ . \ . \ , L_n]$, the average radius of N is given as

$$r_\delta(N) = \inf_\phi e_\delta(\phi, N).$$

An algorithm $\delta_{\delta,N}$ that uses N is optimal iff

$$e_\delta(\phi_{\delta,N}, N) = r_\delta(N).$$

From now on we assume that the functionals forming the information N are μ-normalized, i.e., $\|L_i\|_\mu = 1$, $i = 1, \ldots, n$. This can be done without loss of generality. Indeed, for the information $N = [L_1, \ldots, L_n]$ and an algorithm ϕ, let $N' = [L_1', \ldots, L_n']$ with $L_i' = L_i/\|L_i\|_\mu$, and $\phi'(z') = \phi(z)$, $\forall z' = (z_1', \ldots, z_n')$, $z_i' = z_i/\|L_i\|_\mu$, $i = 1, 2, \ldots, n$. Then $\int_{\mathbb{R}^n} \|Sf - \phi(z)\|^2 \pi_\delta(dz|f, N) = \int_{\mathbb{R}^n} \|Sf - \phi'(z')\|^2 \pi_\delta(dz'|f, N')$, $\forall f \in F$, and $e_\delta(\phi', N') = e_\delta(\phi, N)$. Hence $r_\delta(N) = r_\delta(N')$ and $\phi_{\delta,N'} = \phi_{\delta,N}'$.

We now introduce some notations which are needed to state the theorem.

(a) Let $\langle \circ, \circ \rangle_\mu : F^* \times F^* \to \mathbb{R}$ denote the μ-inner product in F^* which is generated by the correlation operator C_μ of the measure μ,

$$\langle L_1, L_2 \rangle_\mu = L_1(C_\mu L_2) = L_2(C_\mu L_1), \qquad \forall L_1, L_2 \in F^*.$$

(b) Let $\nu = \mu S^{-1}$ be the a priori measure on the space G of solution elements. It is Gaussian with mean zero and the correlation operator C_ν, $C_\nu : G \to G$, such that

$$C_\nu g = S(C_\mu \langle S(\circ), g \rangle), \qquad \forall g \in G.$$

(c) For the information $N = [L_1, \ldots, L_n]$ define the Gram matrix $M_N : \mathbb{R}^n \to \mathbb{R}^n$,

$$M_N = (\langle L_i, L_j \rangle_\mu)_{i,j=1}^n.$$

Let (u_j, η_j) denote the eigenpairs of M_N, $M_N u_j = \eta_j u_j$, $\langle u_i, u_j \rangle_2 = \delta_{i,j}$ (the Kronecker delta), $i, j = 1, \ldots, n$, and $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_m > 0 = \eta_{m+1} = \cdots = \eta_n$. Obviously, $\mathrm{trace}(M_N) = \sum_{i=1}^m \eta_i = \sum_{i=1}^m \langle L_i, L_i \rangle_\mu = n$, since the functionals L_i are μ-normalized.

For $j = 1, 2, \ldots, m$ define the functionals

$$K_j = \eta_j^{-1/2} \langle N(\circ), u_j \rangle_2 = \eta_j^{-1/2} \sum_{i=1}^n u_{i,j} L_i,$$

with $u_j = [u_{j,1}, \ldots, u_{j,n}]^T$. The functionals K_j are μ-orthonormal since

$$\begin{aligned} \langle K_i, K_j \rangle_\mu &= K_i(C_\mu K_j) = \eta_i^{-1/2} \langle N(C_\mu K_j), u_i \rangle_2 \\ &= (\eta_i \eta_j)^{-1/2} \langle N(C_\mu \langle N(\circ), u_j \rangle_2), u_i \rangle_2 \\ &= (\eta_i \eta_j)^{-1/2} \langle M_N u_j, u_i \rangle_2 = \delta_{i,j}. \end{aligned}$$

We also need

LEMMA 3.1. *For any information N we have*

$$N(F) = M_N(\mathbb{R}^n).$$

Proof. We first prove that $N(F) \subset M_N(\mathbb{R}^n)$. Let $f \in C_\mu(F^*)$. Then $f = C_\mu L_f$ for some $L_f \in F^*$. The functional L_f may be represented as $L_f = L' + L''$, where $L' = \sum_{i=1}^n \alpha_i L_i$ and L'' is μ-orthogonal to span$\{L_1, L_2, \ldots, L_n\} \subset F^*$, i.e., $\langle L_i, L''\rangle_\mu = 0$, $\forall\, i = 1, 2, \ldots, n$. Hence $N(C_\mu L'') = 0$ and

$$\begin{aligned} N(f) &= N(C_\mu L_f) = N(C_\mu L') \\ &= N\left(\sum_{i=1}^n \alpha_i C_\mu L_i\right) = \sum_{i=1}^n \alpha_i N(C_\mu L_i) = \left(\sum_{i=1}^n \alpha_i \langle L_j, L_i\rangle_\mu\right)_{i=1}^n \\ &= M_N \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in M_N(\mathbb{R}^n). \end{aligned}$$

Now, let f be an arbitrary element from F. Since the correlation operator C_μ is positive definite then $\overline{C_\mu(F^*)} = F$. Hence $f = \lim_{j\to\infty} f_j$ for some sequence $\{f_j\}_{j=1}^\infty \subset C_\mu(F^*)$. From the fact that $N(f_j) \in M_N(\mathbb{R}^n)$, $\forall j$, and from the continuity of N, we easily conclude that $N(f) \in M_N(\mathbb{R}^n)$.

To show that $M_N(\mathbb{R}^n) \in N(F)$ observe as above that

$$M_N \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = N\left(\sum_{i=1}^n \alpha_i (C_\mu L_i)\right), \qquad \forall [\alpha_1, \alpha_2, \ldots, \alpha_n]^T \in \mathbb{R}^n. \quad \blacksquare$$

We are ready to state the theorem on an optimal algorithm and the average radius of the information N.

THEOREM 3.1. *The optimal algorithm for the information $N = [L_1, L_2, \ldots, L_n]$ is given as*

$$\phi_{\delta,N}(z) = \sum_{i=1}^n y_i S(C_\mu L_i),$$

where $y = [y_1, y_2, \ldots, y_n]^T$ is the solution of the linear system

$$(\delta I + M_N)y = z.$$

The average radius of N is given as

$$r_\delta^2(N) = e_\delta^2(\phi_{\delta,N}, N) = \text{trace}(C_\nu) - \sum_{j=1}^{m} \left(\frac{\eta_j}{\delta + \eta_j}\right) \|SC_\mu K_j\|^2.$$

The theorem needs an explanation for $\delta = 0$ and linearly dependent functionals L_i. In this case, due to Lemma 3.1, the operator $(\delta I + M_N)$ is many-to-one and its range $N(F)$ is a proper subset of $\mathbb{R}^n$. The algorithm $\phi_{\delta,N}$ is then defined only for $z \in N(F)$ and y is taken as the unique vector from $N(F)$ for which the linear system holds.

Proof. From Lemma A1 of the Appendix it follows that for any algorithm ϕ which uses N we have

$$e_\delta^2(\phi, N) = \int_{\mathbb{R}^n} \left(\int_F \|Sf - \phi(z)\|^2 \mu_2(df|z, N)\right) \mu_1(dz; N),$$

where the measure $\mu_2(\circ|z, N)$ is Gaussian with mean $m(z) = \sum_{i=1}^n y_i C_\mu L_i$ (y as in the theorem) and correlation operator $C_{\mu,N}$: $F^* \to F$ given as

$$C_{\mu,N}L = C_\mu L - m(N(C_\mu L)), \qquad \forall L \in F.$$

Note that $C_{\mu,N}$ does not depend on z. Clearly, $\int_F \|Sf - \phi(z)\|^2 \mu_2(df|z, N)$ is minimized for $\phi(z) = Sm(z)$ and thus the optimal algorithm takes the form $\phi_{\delta,N}(z) = Sm(z) = \sum_{i=1}^n y_i S(C_\mu L_i)$. Furthermore, $r_\delta^2(N) = e_\delta^2(\phi_{\delta,N}) = \text{trace}(C_{\nu,N})$, with $C_{\nu,N}$: $G \to G$ being correlation operator of the measure $\mu_2(S^{-1} \circ|z, N)$,

$$C_{\nu,N}g = S(C_{\mu,N}\langle S(\circ), g\rangle), \qquad \forall g \in G.$$

Observe that $N(C_\mu L) = \sum_{j=1}^n \langle N(C_\mu L), u_j\rangle_2 u_j = \sum_{j=1}^n \eta_j^{1/2} K_j(C_\mu L)u_j$ and $m(u_j) = \sum_{i=1}^n ((\delta I + M_N)^{-1}u_j)_i C_\mu L_i = (\delta + \eta_j)^{-1} \sum_{i=1}^n u_{ji} C_\mu L_i = \eta_j^{1/2}/(\delta + \eta_j) C_\mu L_i$. Hence $m(NC_\mu L) = \sum_{j=1}^n \eta_j^{1/2} K_j(C_\mu L) m(u_j) = \sum_{j=1}^m \eta_j/(\delta + \eta_j) L(C_\mu K_j) C_\mu K_j$. Furthermore,

$$C_{\nu,N}g = C_\nu g - \sum_{j=1}^{m} \eta_j/(\delta + \eta_j)\langle S(C_\mu K_j), g\rangle S(C_\mu K_j).$$

For an orthonormal basis $\{g_i\}_{i=1}^\infty$ in G we have

$$\begin{aligned}\operatorname{trace}(C_{\nu,N}) &= \sum_{i=1}^{\infty} \langle C_{\nu,N} g_i, g_i \rangle \\ &= \sum_{i=1}^{\infty} \langle C_\nu g_i, g_i \rangle - \sum_{i=1}^{\infty} \sum_{j=1}^{m} \eta_j/(\delta + \eta_j) \langle SC_\mu K_j, g_i \rangle^2 \\ &= \operatorname{trace}(C_\nu) - \sum_{j=1}^{m} \eta_j/(\delta + \eta_j) \sum_{i=1}^{\infty} \langle SC_\mu K_j, g_i \rangle^2 \\ &= \operatorname{trace}(C_\nu) - \sum_{j=1}^{m} \eta_j/(\delta + \eta_j) \| SC_\mu K_j \|^2,\end{aligned}$$

as claimed. ∎

From Theorem 3.1 we can easily compare the average radii for positive δ and for $\delta = 0$.

COROLLARY 3.1.

$$r_\delta^2(N) = r_0^2(N) + \delta \cdot \sum_{j=1}^{m} \frac{\|SC_\mu K_j\|^2}{\delta + \eta_j}.$$

We now illustrate Theorem 3.1 by assuming that the information N consists of the repetitive evaluations of some functionals.

EXAMPLE 3.1. Let L_i, $i = 1, 2, \ldots, s$, be μ-orthonormal. $\langle L_i, L_j \rangle_\mu = \delta_{ij}$. Consider the information operator

$$N = [\underbrace{L_1, \ldots, L_1}_{k_1}, \underbrace{L_2, \ldots, L_2}_{k_2}, \ldots, \underbrace{L_s, \ldots, L_s}_{k_s}],$$

where $k_1 \geq k_2 \geq \cdots \geq k_s \geq 1$. Applying Theorem 3.1 we get that the radius of the information N is given by

$$r_\delta^2(N) = \operatorname{trace}(C_\nu) - \sum_{i=1}^{s} k_i/(\delta + k_i) \|SC_\mu L_i\|^2.$$

The optimal algorithm for N takes the form

$$\phi_{\delta,N}(z) = \sum_{i=1}^{s} \left((\delta + k_i)^{-1} \sum_{j=1}^{k_i} z_{ij} \right) S(C_\mu L_i),$$

where $z = [z_{11}, \ldots, z_{1k_1}, \ldots, z_{s1}, \ldots, z_{sk_s}]^T \in \mathbb{R}^{k_1 + \cdots + k_s}$.

In particular, for $\delta = 0$ and $k_1 = k_2 = \cdots = k_s = 1$ we get the known formulas for the exact information case; see IBC.

4. nTH OPTIMAL RADIUS AND INFORMATION

Denote by $\mathcal{N}_n$ the class of all information operators each of which consists of n continuous linear functionals. Define the nth optimal average radius as

$$r_\delta(n) = \inf_{N \in \mathcal{N}_n} r_\delta(N).$$

An information operator $N_{\delta,n}$ is the nth optimal iff

$$r_\delta(N_{\delta,n}) = r_\delta(n).$$

Let $C_\nu\colon G \to G$ be, as in Section 3, the correlation operator of the measure $\nu = \mu S^{-1}$. Let $\lambda_1 \geq \lambda_2 \geq \ldots$ be the dominating eigenvalues of C_ν, and let $\xi_1, \xi_2, \ldots$ be the orthonormal in G basis of the corresponding eigenvectors. If dim $G < +\infty$ then we formally set $\lambda_i = 0$ for $i >$ dim G. Set $K_i^* = \lambda_i^{-1/2}\langle S(\circ), \xi_i\rangle$, $i = 1, 2, \ldots$. The functionals K_i^* are μ-orthonormal. It is known that the information $[K_1^*, \ldots, K_n^*]$ is nth optimal for the exact information, $\delta = 0$, see IBC. We show that a proper linear combinations of K_i^*'s form the nth optimal information for noisy information, $\delta > 0$. To show this we first need the following lemma.

LEMMA 4.1. *For any nonnegative $\eta_1, \eta_2, \ldots, \eta_n$, $\sum_{i=1}^n \eta_i = n$, there exists a matrix $W = \{w_{ij}\}_{ij=1}^n$ such that*

$$\sum_{j=1}^n w_{ij}^2 = 1, \qquad \forall i = 1, 2 \ldots, n,$$

and

$$\sum_{s=1}^n w_{si}w_{sj} = \begin{cases} 0 & i \neq j, \\ \eta_i & i = j, \end{cases} \qquad \forall i, j = 1, 2, \ldots, n.$$

Proof. We prove the lemma by constructing the matrix W inductively on n.

For $n = 1$ we have $\eta_1 = 1$ and $w_{11} = 1$. Suppose that $n > 1$. Assume, without loss of generality, that $\eta_{n-1} = \max_i \eta_i$, $\eta_n = \min_i \eta_i$. If $\eta_n = 1$ or $\eta_{n-1} = 1$ then $\eta_i = 1$, $\forall i$, and W can be taken as the identity matrix.

Assume that $\eta_n < 1 < \eta_{n-1}$. Set $\eta = \eta_{n-1} + \eta_n - 1 > 0$,

$$a = \left(\frac{\eta_{n-1}(\eta_{n-1} - 1)}{\eta(\eta_{n-1} - \eta_n)}\right)^{1/2}, \qquad b = (1 - a^2)^{1/2},$$

$$c = \left(\frac{\eta_{n-1}(1 - \eta_n)}{(\eta_{n-1} - \eta_n)}\right)^{1/2}, \qquad d = -(1 - c^2)^{1/2}.$$

Note that the above values are well defined. Let $W_{n-1} = (w_1 \ w_2 \ \ldots \ w_{n-1})$, $w_i \in \mathbb{R}^{n-1}$, $i = 1, 2, \ldots, n-1$, be the required matrix for $\eta_1, \eta_2, \ldots, \eta_{n-2}, \eta$. Direct calculations yield that the lemma holds with the matrix W_n,

$$W_n = \begin{pmatrix} w_1 & w_2 & \ldots & w_{n-2} & aw_{n-1} & bw_{n-1} \\ 0 & 0 & \ldots & 0 & c & d \end{pmatrix}. \quad \blacksquare$$

We are now ready to state the theorem on the nth optimal average radius and information.

THEOREM 4.1. *The nth optimal average radius is*

$$r_\delta^2(n) = \delta \cdot \frac{\left(\sum_{j=1}^{k} \lambda_j^{1/2}\right)^2}{n + \delta k} + \sum_{j=k+1}^{\infty} \lambda_j,$$

where $k = k(\delta, n)$ is the greatest integer no larger than n such that

$$\delta \cdot \frac{\left(\sum_{i=1}^{k} \lambda_j^{1/2}\right)}{n + \delta k} \le \lambda_k^{1/2}.$$

The nth optimal information is

$$N_{\delta,N} = [L_1^*, L_2^*, \ldots, L_n^*],$$

where $L_i^* = \sum_{j=1}^k w_{ij} K_j^*$, $i = 1, 2, \ldots, n$, and $W = \{w_{ij}\}_{ij=1}^n$ is the matrix from Lemma 4.1 applied for

$$\eta_i^* = \begin{cases} \dfrac{n + \delta k}{\left(\sum_{j=1}^{k} \lambda_j^{1/2}\right)} \cdot \lambda_i^{1/2} - \delta, & i = 1, 2, \ldots, k, \\ 0, & i = k+1, \ldots, n. \end{cases}$$

Proof. For any $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m \geq 0$ and $a_i, b_i, i = 1, \ldots, m$, such that $\sum_{i=1}^s a_i \leq \sum_{i=1}^s b_i$, $\forall s = 1, 2, \ldots, m$, we have $\sum_{i=1}^m \beta_i a_i \leq \sum_{i=1}^m \beta_i b_i$. Indeed, using induction on m we have

$$\sum_{i=1}^m \beta_i a_i = \beta_m \left(\sum_{i=1}^m a_i\right) + \sum_{i=1}^{m-1} (\beta_i - \beta_m) a_i \leq \beta_m \left(\sum_{i=1}^m b_i\right)$$

$$+ \sum_{i=1}^{m-1} (\beta_i - \beta_m) b_i = \sum_{i=1}^m \beta_i b_i.$$

Define $a_j = \|SC_\mu K_j\|^2$, $b_j = \lambda_j$, $j = 1, \ldots, n$. It is known that $\sum_{j=1}^s a_j = \sum_{j=1}^s \|SC_\mu K_j\|^2 \leq \sum_{j=1}^s \lambda_j = \sum_{j=1}^s b_j$. From Theorem 3.1 and the above inequality we get, for $N \in \mathcal{N}_n$ and $\beta_j = \eta_j/(\delta + \eta_j)$,

$$\begin{aligned} r_\delta^2(N) &= \operatorname{trace}(C_\nu) - \sum_{j=1}^m \eta_j/(\delta + \eta_j)\|SC_\mu K_j\|^2 \\ &\geq \operatorname{trace}(C_\nu) - \sum_{j=1}^m \eta_j/(\delta + \eta_j)\lambda_j = \sum_{j=1}^\infty \lambda_j - \sum_{j=1}^m \eta_j/(\delta + \eta_j)\lambda_j \\ &= \delta \sum_{j=1}^m \eta_j/(\delta + \eta_j) + \sum_{j=m+1}^\infty \lambda_j =: \Omega(\eta_1, \eta_2, \ldots, \eta_m). \end{aligned}$$

Hence $r_\delta^2(N) \geq \min \Omega(\eta_1, \ldots, \eta_m)$, where the minimum is taken over all $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_m > 0$, $m \leq n$, such that $\sum_{i=1}^m \eta_i = n$. Using the standard technique we get that the function Ω has its minimum at the point $\eta^* = (\eta_1^*, \ldots, \eta_k^*)$, with η_i^* and k as in the theorem, and

$$r_\delta^2(n) \geq \Omega(\eta_1^*, \ldots, \eta_k^*) = \delta \frac{\left(\sum_{j=1}^k \lambda_j^{1/2}\right)^2}{n + \delta k} + \sum_{j=k+1}^\infty \lambda_j.$$

To complete the proof it is enough to show that $r_\delta^2(N_{\delta,n}) = \Omega(\eta_1^*, \ldots, \eta_k^*)$. Indeed, for $z_i = \eta_i^{-1/2}(w_{1i}, \ldots, w_{ni})^T$, $i = 1, 2, \ldots, k$, and for the matrix $M = (\langle L_i^*, L_j^*\rangle_\mu)_{ij=1}^n$ we have $\|z\|_2 = 1$ and $(Mz_i)_s = (\eta_i^*)^{-1/2} \sum_{j=1}^n \langle L_s, L_j\rangle_\mu w_{ij} = (\eta_i^*)^{-1/2} \sum_{j=1}^n (\sum_{r=1}^n w_{sr} w_{jr}) w_{ji} = (\eta_i^*)^{-1/2} \sum_{r=1}^n (\sum_{j=1}^n w_{jr} w_{ji}) w_{sr} = (\eta_i^*)^{-1/2} \eta_i^* w_{si} = \eta_i^* z_{is}$, $\forall s = 1, 2, \ldots, n$. Therefore $Mz_i = \eta_i^* z_i$ $\forall i = 1, 2, \ldots, k$, and the μ-orthonormal functionals K_i, $i = 1, 2, \ldots, k$, corresponding to the information $N_{\delta,n}$ are given by

$$K_i = (\eta_i^*)^{-1/2}\langle N_{\delta,n}(\circ), z_i\rangle_2 = (\eta_i^*)^{-1/2} \sum_{j=1}^{n} L_j^* z_{ij}$$

$$= (\eta_i^*)^{-1/2} \sum_{j=1}^{n} \left(\sum_{s=1}^{n} w_{js} K_s\right) z_{ij} = (\eta_i^*)^{-1/2} \sum_{s=1}^{n} \left(\sum_{j=1}^{n} w_{js} z_{ij}\right) K_s^*$$

$$= (\eta_i)^{-1} \sum_{s=1}^{n} \left(\sum_{j=1}^{n} w_{js} w_{ji}\right) K_s^* = K_i^*.$$

From this we conclude that

$$r_\delta^2(N_{\delta,n}) = \mathrm{trace}(C_\nu) - \sum_{j=1}^{k} \eta_j^*/(\delta + \eta_j^*)\|SC_\mu K_j^*\|^2$$

$$= \delta \sum_{j=1}^{k} \lambda_j/(\delta + \eta_j^*) + \sum_{j=k+1}^{\infty} \lambda_j = \Omega(\eta_1^*, \ldots, \eta_k^*),$$

as claimed. ∎

Theorem 4.1, together with Lemma 3.1, shows how to obtain the nth optimal information for given $\delta \geq 0$ and the eigenpairs λ_i, ξ_i, $i = 1, 2, \ldots, n$, of the correlation operator C_ν. Observe that the cost of computing $\eta_1^*, \ldots, \eta_n^*$ plus the cost of finding the coefficients w_{ij}, $i, j = 1, 2, \ldots, n$, of the matrix W is proportional to n^2 arithmetic operations and comparisons. Therefore there are no difficulties in obtaining the optimal functionals $L_1^*, L_2^*, \ldots, L_n^*$.

Note that the nth optimal information is, in general, not uniquely determined. For instance, for $\delta = 0$, any information $N = [L_1, L_2, \ldots, L_n]$ with the functionals L_i such that $\mathrm{span}\{L_1, \ldots, L_n\} = \mathrm{span}\{K_1^*, \ldots, K_n^*\}$ is nth optimal. In the general case, i.e., for $\delta \geq 0$, we always have that $\mathrm{span}\{L_1^*, \ldots, L_n^*\} \subseteq \mathrm{span}\{K_1^*, \ldots, K_k^*\}$, with k as in the theorem.

EXAMPLE 4.1. We show the nth optimal information for approximation in the two-dimensional space. That is, $F = G = \mathbb{R}^2$, $Sf = f$, $\forall f \in F$. Let $C_\mu e_i = \lambda_i e_i$, $\lambda_1 \geq \lambda_2 > 0$, where e_i is the ith versor, $i = 1, 2$. Let $e_i' = \lambda_i^{-1/2} e_i$. For $\lambda_1/\lambda_2 > (n/\delta + 1)^2$ we have $k = k(\delta, n) = 1$ and

$$N_{\delta,n} = [\underbrace{\langle \circ, e_1'\rangle, \ldots, \langle \circ, e_1'\rangle}_{n}].$$

Let $\lambda_1/\lambda_2 \leq (n/\delta + 1)^2$. Then $k = k(\delta, n) = 2$ and $\eta_i^* = ((n + 2\delta)/(\lambda_1^{1/2} + \lambda_2^{1/2}))\lambda_i^{1/2} - \delta$, $i = 1, 2$. Let $\eta_i^* = s_i + \alpha_i$, with $s_i = \lfloor \eta_i^* \rfloor$, $0 \leq \alpha_i < 1$. Clearly,

$\alpha_1 + \alpha_2 = 1$ and $s_1 + s_2 = n - 1$. The nth optimal information $N_{\delta,n}$ given by Theorem 4.1 is now of the form

$$N_{\delta,n} = [\underbrace{\langle \circ, e_1' \rangle, \ldots, \langle \circ, e_1' \rangle}_{s_1 - 1}, \underbrace{\langle \circ, e_2' \rangle, \ldots, \langle \circ, e_2' \rangle}_{s_2 - 1}, \langle \circ, e \rangle, \langle \circ, x_1 \rangle, \langle \circ, x_2 \rangle],$$

where $e = e_1'$ if $\alpha_1 \geq \alpha_2$ and $e = e_2'$ otherwise, and

$$x_1 = \left[\left(\frac{2-\alpha}{2\lambda_1}\right)^{1/2}, \left(\frac{\alpha}{2\lambda_2}\right)^{1/2}\right]^T, \qquad x_2 = \left[\left(\frac{2-\alpha}{2\lambda_1}\right)^{1/2}, -\left(\frac{\alpha}{2\lambda_2}\right)^{1/2}\right]^T,$$

with $\alpha = \max\{\alpha_1, \alpha_2\}$. In particular, for $n = 2$ and $\lambda_1/\lambda_2 \leq (2/\delta + 1)^2$ we have $\eta_i^* = (2(1 + \delta)/(\lambda_1^{1/2} + \lambda_2^{1/2}))\lambda_i^{1/2} - \delta$, $i = 1, 2$, and $N_{\delta,2} = [\langle \circ, y_1 \rangle, \langle \circ, y_2 \rangle]$, where

$$y_1 = \left[\left(\frac{\eta_1^*}{2\lambda_1}\right)^{1/2}, \left(\frac{\eta_2^*}{2\lambda_2}\right)^{1/2}\right]^T, \qquad y_2 = \left[\left(\frac{\eta_1^*}{2\lambda_1}\right)^{1/2}, -\left(\frac{\eta_2^*}{2\lambda_2}\right)^{1/2}\right]^T.$$

Note that $\langle y_1, y_2 \rangle_\mu = (1 + \delta)((\lambda_1^{1/2} - \lambda_2^{1/2})/(\lambda_1^{1/2} + \lambda_2^{1/2}))$.

We now comment on the nth optimal radius. It is easy to see that for fixed n we have

$$\lim_{\delta \to 0^+} r_\delta(n) = \sqrt{\sum_{j=n+1}^{\infty} \lambda_j} = r_0(n),$$

$$\lim_{\delta \to \infty} r_\delta(n) = \sqrt{\sum_{j=1}^{\infty} \lambda_j} = \sqrt{\mathrm{trace}(C_\mu)} = r_\delta(0).$$

Now, let δ be fixed and let $n \to +\infty$. For $\dim S(F) = +\infty$ we have $k(\delta, n) \to +\infty$ and

$$r_\delta^2(n) \leq \lambda_k^{1/2} \left(\sum_{j=1}^{k} \lambda_j^{1/2}\right) + \sum_{j=k+1}^{\infty} \lambda_j = \sum_{j=1}^{k} (\lambda_j \lambda_k)^{1/2} + \sum_{j=k+1}^{\infty} \lambda_j \to 0.$$

Obviously, for $\dim S(F) < +\infty$ we also have $r^2(n) \to 0$, as $\delta \to 0^+$. On the other hand, $r_\delta^2(n) = \Omega(\eta_1^*, \ldots, \eta_k^*) \geq \delta\lambda_1/(\delta + \eta_1^*) \geq \delta\lambda_1/(\delta + n)$ and this inequality is sharp for $S \in F^*$. Hence, for $\lambda_1 > 0$ (which holds for $S \neq 0$) and $\delta > 0$ the sequence of the nth optimal radii never tends to zero faster than $n^{-1/2}$.

EXAMPLE 4.2. We illustrate the dependence of $r_\delta(n)$ on δ and $\{\lambda_j\}$ for $\lambda_j = (aj)^{-p}$, $j = 1, 2, \ldots$, where $p > 1$. We have

$$r_\delta^2(n) = \rho(\delta, n) \cdot (1 + o(1)), \qquad \text{as } n \to +\infty,$$

where the function $\rho(\delta, n)$ is given below ($\delta \geq 0$):

$$p > 2: \quad \frac{\delta Z_{a,p}^2}{n} + \frac{1}{a^p(p-1)n^{p-1}}, \; Z_{a,p} = \sum_{j=1}^{\infty} (aj)^{-p}$$

$$p = 2: \quad \frac{\delta \ln^2 n}{a^2 n} + \frac{1}{an}$$

$$1 < p < 2, \; \delta \leq (2-p)/p: \quad \frac{1}{n^{p-1}a^p}\left(\frac{\delta}{1+\delta}\frac{1}{(1-p/2)^2} + \frac{1}{p-1}\right)$$

Remark 4.1. Let $\bar{\mathcal{N}}_n$ be the class of all information operators $N = [L_1, L_2, \ldots, L_n]$ such that $L_i \in \{K_1^*, K_2^*, K_3^*, \ldots\}$, $\forall i$. Information $N \in \mathcal{N}_n^*$ has some advantages. It is easy to obtain and the optimal algorithm, given in Example 3.1, takes a very simple form. Therefore it is interesting to see how much can be lost by such a restriction of permissible functionals. Define

$$\bar{r}_\delta(n) = \inf_{N \in \bar{\mathcal{N}}_n} r_\delta(N).$$

If $\delta = 0$ then we know that $\bar{r}_\delta(n) = r_\delta(n)$. To find $\bar{r}_\delta(n)$ for $\delta > 0$ we must, however, evaluate the maximum of $\sum_{j=1}^n \lambda_j/(s_i + \delta)$ over all integers $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$ such that $\sum_{j=1}^n s_j = n$ (this follows from Example 3.1). Furthermore, for the optimal values s_j^*, $j = 1, 2, \ldots, n$, we have $\bar{r}_\delta^2(n) = \text{trace}(C_\nu) - \sum_{j=1}^n \lambda_j/(s_j^* + \delta)$, and $\bar{r}_\delta^2(n) = \bar{r}_\delta^2(\bar{N}_{\delta,n})$, where

$$\bar{N}_{\delta,n} = [\underbrace{K_1^*, \ldots, K_1^*}_{s_1^*}, \ldots, \underbrace{K_n^*, \ldots, K_n^*}_{s_n^*}].$$

We can, however, easily show an upper bound on $\bar{r}_\delta(n)$ in terms of the optimal radii. To this end, let $m_j^* = \lceil \eta_j^* \rceil$, $j = 1, \ldots, n$, where η_j^* are the optimal eigenvalues for n from Theorem 3.1. Define

$$N = [\underbrace{K_1^*, \ldots, K_1^*}_{m_1^*}, \ldots, \underbrace{K_n^*, \ldots, K_n^*}_{m_n^*}].$$

Clearly, N consists of at most $n + k(\delta, n) - 1 \leq 2n - 1$ functionals and $\bar{r}_\delta(N) \leq r_\delta(n)$. From this we conclude that $\bar{r}_\delta(2n-1) \leq r_\delta(n)$, or

$$r_\delta(n) \leq \bar{r}_\delta(n) \leq r_\delta(\lceil n/2 \rceil), \qquad \forall n = 1, 2, \ldots.$$

It is also easy to show that for $\lambda_j = (aj)^{-p}$, $p > 1$, we have

$$r_\delta(n) \leq r_\delta(n) \leq 2^q r_\delta(n)(1 + o(1)) \qquad \text{as } n \to +\infty,$$

where $q = \min\{1/2, (p - 1)/2\}$.

5. Adaptive Noisy Information

Until now we have analyzed nonadaptive information. That is, for any information operator $N = [L_1, L_2, \ldots, L_n]$ the functionals L_i were given simultaneously and they did not change during the successive observations of the values $z_i = L_i(f) + x_i$, $i = 1, 2, \ldots$. In this and in the next sections we deal with more general class of adaptive information. That is, the functionals L_i as well as the total number of them may vary, depending on the observed z_i, $i = 1, 2, \ldots$. More precisely, the exact ($\delta = 0$) adaptive information N^a is of the form

$$N^a(f) = [L_1(f), L_2(f; y_1), \ldots, L_{n(y)}(f; y_1, \ldots, y_{n(y)-1})],$$

where $y_i = L_i(f; y_1, \ldots, y_{i-1})$ and $L_i(\circ; y_1, \ldots, y_{i-1})$, $i = 1, 2, \ldots, n(y)$, are continuous linear functionals for any fixed $y_1, y_2, \ldots, y_{i-1}$. The number $n(y)$ is determined by the so-called termination functions $\mathrm{ter}_i\colon \mathbb{R}^i \to \{0, 1\}$, $i = 1, 2, \ldots$,

$$n(y) = \min\{i\colon \mathrm{ter}_i(y_1, y_2, \ldots, y_i) = 1\}.$$

We assume for simplicity that $n(y) < +\infty$, $\forall y$. We now turn to noisy adaptive information. This means that instead of the exact value $y = N(f)$ we observe $z = [z_1, z_2, \ldots, z_{n(z)}]$, where

$$z_i = L_i(f; z_1, \ldots, z_{i-1}) + x_i, \qquad i = 1, 2, \ldots, n(z),$$

$n(z) = \min\{i\colon \mathrm{ter}_i(z_1, \ldots, z_i) = 1\}$, and the vector $x = [x_1, x_2, \ldots, x_{n(z)}]$ denotes the noise.

We stress that the choice of the functionals L_i as well as the total number of them $n(z)$ is now based on the observed values $z_1, z_2, \ldots$, and not on the exact $L_i(f)^1s$. Hence, $n(z)$ may vary even for fixed $f \in F$.

Define the sets

$$\begin{aligned} B_m = \{[z_1, z_2, \ldots, z_m] \in \mathbb{R}^m : {} & \mathrm{ter}_m(z_1, \ldots, z_m) = 1, \\ & \mathrm{ter}_i(z_1, \ldots, z_i) = 0, \forall i = 1, 2, \ldots, m - 1\}. \end{aligned}$$

Let $Y = \bigcup_{m=1}^{\infty} B_m$. Then the observed value z always belongs to Y. Assuming that the functionals $L_i(f; \circ, \ldots, \circ)$ are Borel measurable and that B_i, $i = 1, 2, \ldots$, are Borel sets in $\mathbb{R}^i$, we have for fixed $f \in F$

$$\text{Prob}(z \in B) = \pi_\delta(B \mid f, N^a), \qquad \forall B \in \mathbb{B}(Y),$$

where $\mathbb{B}(Y) = \{C \subset Y : C \cap B_m \in \mathbb{B}(\mathbb{R}^m), \forall m\}$.

For $\delta > 0$ the measure π_δ is given by

$$\pi_\delta(B \mid f, N^a) = \sum_{m=1}^{\infty} (2\pi)^{-m/2} \int_{B \cap B_m} (\sigma_1 \sigma_2(t_1) \ldots \sigma_m(t_1, \ldots, t_{m-1}))^{-1}$$
$$\exp\left\{-\frac{1}{2}\sum_{i=1}^{m} \frac{(t_i - L_i(f; t_1, \ldots, t_{i-1}))^2}{\sigma^2(t_1, \ldots, t_{i-1})}\right\}$$
$$dt_m dt_{m-1} \ldots dt_1,$$

where $\sigma_i^2(t_1, \ldots, t_{i-1}) = \delta \cdot \|L_i(\circ; t_1, \ldots, t_{i-1})\|_\mu^2$. For $\delta = 0$ we have

$$\pi_\delta(B \mid f, N^a) = \begin{cases} 1 & \text{if } N^a(f) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

An algorithm ϕ that uses adaptive information N^a is now defined on the set Y; i.e., ϕ: $Y \to G$. The error of ϕ is equal to

$$e_\delta(\phi, N^a) = \left\{\int_F \left(\int_Y \|Sf - \phi(z)\|^2 \pi_\delta(dz \mid f, N^a)\right) \mu(df)\right\}^{1/2}$$
$$= \left\{\sum_{m=1}^{\infty} \int_F \left(\int_{B_m} \|Sf - \phi(z)\|^2 \pi_\delta(dz \mid f, N^a)\right) \mu(df)\right\}^{1/2}.$$

The average radius of N^a and the optimal algorithm ϕ_{δ,N^a} are defined as in the nonadaptive case; i.e.,

$$r_\delta(N^a) = \inf_\phi e_\delta(\phi, N^a),$$

and ϕ_{δ,N^a} is optimal for N^a iff

$$e(\phi_{\delta,N^a}, N^a) = r_\delta(N^a).$$

We study whether adaptive information may be more powerful than nonadaptive.

6. Adaptive Information versus Nonadaptive

For adaptive information N^{a}, let $\mathrm{card}_\delta(N^{\mathrm{a}})$ denote the average number of functional evaluations when N^{a} is used. That is,

$$\mathrm{card}_\delta(N^{\mathrm{a}}) = \int_F \left(\int_Y n(z)\pi_\delta(dz|f, N^{\mathrm{a}})\right) \mu(df)$$

$$= \sum_{m=1}^{\infty} \int_F m \cdot \pi_\delta(B_m|f, N^{\mathrm{a}})\mu(df).$$

Obviously, $\mathrm{card}_\delta(N^{\mathrm{a}}) = n$ whenever $n(z) \equiv n$; i.e., $Y = \mathbb{R}^n$. For fixed $z = [z_1, z_2, \ldots, z_m] \in Y$, let N_z^{non} be the nonadaptive information of the form

$$N_z^{\mathrm{non}} = [L_1(\circ), L_2(\circ; z_1), \ldots, L_m(\circ; z_1, \ldots, z_{m-1})].$$

As in Section 3 we assume, without loss of generality, that $\|L_i(\circ; z_1, \ldots, z_{i-1})\|_\mu = 1$, $\forall i = 1, 2, \ldots, \forall z \in Y$.

We now state the first theorem of this section.

THEOREM 6.1. *Let N^{a} be adaptive information with* $\mathrm{card}_\delta(N^{\mathrm{a}}) < +\infty$. *Then there exist real a^* and $z', z'' \in Y$ such that for the information*

$$N^*(f) = \begin{cases} N_{z'}^{\mathrm{non}}(f) & \text{if } z_1 = L_1(f) + x_1 \le a^*, \\ N_{z''}^{\mathrm{non}}(f) & \text{otherwise,} \end{cases}$$

we have

$$\mathrm{card}_\delta(N^*) \le \mathrm{card}_\delta(N^{\mathrm{a}}) \quad \textit{and} \quad r_\delta(N^*) \le r_\delta(N^{\mathrm{a}}).$$

Proof. From Lemma A2 of the Appendix it follows that for any algorithm ϕ which uses N^{a} we have

$$e_\delta^2(\phi, N^{\mathrm{a}}) = \sum_{m=1}^{\infty} \int_{B_m} \left(\int_F \|Sf - \phi(z)\|^2 \mu(df|z, N_z^{\mathrm{non}})\right) \mu_1(dz; N^{\mathrm{a}}).$$

Hence, we can apply the technique used by Wasilkowski (1986); see also the proof of Theorem 5.6.2 from IBC (Chap. 6) to conclude that there exist two indices $m_1 \le m_2$ and real a, $0 \le a \le 1$, such $am_1 + (1 - a)m_2 \le \mathrm{card}_\delta(N^{\mathrm{a}})$ and

$$r_\delta^2(N^{\mathrm{a}}) \ge a r_\delta^2(N_{z_{m_1}}^{\mathrm{non}}) + (1 - a) r_\delta^2(N_{z_{m_2}}^{\mathrm{non}}).$$

Thus the theorem holds for $z' = z_{m_1}$, $z'' = z_{m_2}$, and a^* such that $\omega((-\infty, a^*)) = a$, where

$$\begin{aligned} \omega(B) &= \mathrm{Prob}(z_1 \in B) \\ &= (2\pi(1 + \delta))^{-1/2} \int_B \exp\left\{\frac{t^2}{2(1 + \delta)}\right\} dt, \qquad \forall B \in \mathbb{B}(\mathbb{R}). \quad \blacksquare \end{aligned}$$

We know that for information $N_{z_{m_i}}^{\mathrm{non}}$, $i = 1, 2$, from the proof we have $r_\delta^2(N_{z_{m_i}}^{\mathrm{non}}) \ge \delta\lambda_1/(\delta + m_i)$, where λ_1 is the dominating eigenvalue of the correlation operation C_ν. Furthermore,

$$\begin{aligned} r_\delta^2(N^{\mathrm{a}}) \ge r_\delta^2(N^*) &= a r_\delta^2(N_{z_{m_1}}^{\mathrm{non}}) + (1 - a) r_\delta^2(N_{z_{m_2}}^{\mathrm{non}}) \\ &\ge \delta\lambda_1(a/(\delta + m_1) + (1 - a)/(\delta + m_2)) \\ &\ge \delta\lambda_1/(a(\delta + m_1) + (1 - a)(\delta + m_2)) \\ &= \delta\lambda_1/(\delta + \mathrm{card}_\delta(N^*)) \ge \delta\lambda_1/(\delta + \mathrm{card}_\delta(N^{\mathrm{a}})). \end{aligned}$$

Hence,

COROLLARY 6.1. *For any adaptive information N^{a} we have*

$$r_\delta^2(N^{\mathrm{a}}) \ge \frac{\delta\lambda_1}{\delta + \mathrm{card}_\delta(N^{\mathrm{a}})}.$$

We now show an important lemma, from which it follows that adaptive information is not more powerful than nonadaptive.

LEMMA 6.1. *The sequence $r_\delta^2(n)$ is convex; i.e.,*

$$(r_\delta^2(n - 1) + r_\delta^2(n + 1))/2 \ge r_\delta^2(n), \qquad \forall n \ge 2.$$

Proof. We show the equivalent inequality $r_\delta^2(n - 1) - r_\delta^2(n) \ge r_\delta^2(n) - r_\delta^2(n + 1)$, $\forall n \ge 2$. To this end, let $k = k(\delta, n)$ and $l = k(\delta, n + 1)$. Consider the difference

$$\Delta_1 = r_\delta^2(n) - r_\delta^2(n+1)$$
$$= \delta/(n+\delta k) \cdot \left(\sum_{i=1}^{k} \lambda_i^{1/2}\right)^2 + \sum_{i=k+1}^{l} \lambda_i - \delta/(n+1+\delta l) \cdot \left(\sum_{i=1}^{l} \lambda_i^{1/2}\right)^2,$$

as a function of $\lambda_j^{1/2}$, $j = k+1, \ldots, n, n+1$. Then

$$\frac{\partial \Delta_1}{\partial \lambda_j^{1/2}} = \begin{cases} 0, & \text{for } j = l+1, \ldots, n+1, \\ 2\left(\lambda_j^{1/2} - \delta/(n+1+\delta l) \sum_{i=1}^{l} \lambda_i^{1/2}\right), & \text{for } j = k+1, \ldots, l. \end{cases}$$

From the definition of l and from the inequality $\lambda_j^{1/2} \ge \lambda_l^{1/2}$, $\forall j \le l$, it follows that the derivative $\partial\Delta_1/\partial\lambda_j^{1/2}$ is nonnegative. Hence Δ_1 is a nondecreasing function of λ_j, $j > k$, if only the parameter $k = k(\delta, n)$ is constant. On the other hand, this does not affect $\Delta_0 = r_\delta^2(n-1) - r_\delta^2(n)$, since it is independent of λ_j, $j > k$. Let

$$a = \begin{cases} \lambda_n^{1/2}, & \text{for } k = n, \\ \delta/(n+\delta k)\left(\sum_{i=1}^{k} \lambda_i^{1/2}\right), & \text{for } k < n. \end{cases}$$

It is easy to check that if $\alpha_{k+1} \ge \cdots \ge \alpha_n$ and $\lambda_i \le \alpha_i < a^2$, $i = k+1, \ldots, n$, then $\delta/(n+\delta s) \cdot (\sum_{i=1}^{k} \lambda_i^{1/2} + \sum_{i=k+1}^{s} \alpha_i^{1/2}) \ge \delta/(n+\delta s) \cdot (\sum_{i=1}^{k} \lambda_i^{1/2} + (s-k)\alpha_s^{1/2}) > \alpha_s^{1/2}$, $\forall s$, $k < s \le n$. Furthermore, $(\delta/(n+\delta s) \cdot (\sum_{i=1}^{k} \lambda_i^{1/2} + \sum_{i=k+1}^{s} a) = a$, $\forall s > k$. From this and from the continuity of $r_\delta(n)$ with respect to the eigenvalues it follows that for the sequence $\{\lambda_i'\}$,

$$\lambda_i' = \begin{cases} \lambda_i, & i = 1, 2, \ldots, k, \\ a^2, & i = k+1, \ldots, n+1, \\ 0, & i = n+2, \ldots, \end{cases}$$

and for the corresponding Δ_0', Δ_1', we have $\Delta_0' = \Delta_0$ and $\Delta_1' \ge \Delta_1$. To complete the proof it is enough to show that $\Delta_0' \ge \Delta_1'$. Indeed, for $A = \sum_{i=1}^{n-1} (\lambda_i')^{1/2}$ we have

$$r_\delta'(n+1)^2 = \delta/((1+\delta)(n+1)) \cdot (A+2a)^2,$$
$$r_\delta'(n)^2 = \delta/((1+\delta)n) \cdot (A+a)^2 + a^2,$$

$$r'_\delta(n-1)^2 \geq \inf\left\{\Omega(\eta_1, \ldots, \eta_{n-1}) \colon \sum_{i=1}^{n-1} \eta_i = n-1\right\}$$
$$= \delta/((1+\delta)(n-1)) \cdot A^2 + 2a^2$$

(Ω as in the proof of Theorem 4.1.). Direct calculations show that the claimed inequality holds. Moreover, $\Delta_0 = \Delta_1$ holds only in two cases: $\delta = 0$, $\lambda_n = \lambda_{n+1}$ or $\delta > 0$, $\lambda_1 = \lambda_2 = \cdots = \lambda_{n+1}$. ∎

As in IBC for exact information (see Theorem 5.6.3 of Chap. 6) we conclude the following

THEOREM 6.2. *Let N^a be any adaptive information with finite cardinality. Let $m = \lceil \mathrm{card}_\delta(N^a) \rceil$. Then for the nth optimal nonadaptive information $N_{\delta,m}$, we have*

$$r_\delta(N_{\delta,m}) \leq r_\delta(N^a).$$

7. Complexity Results

In this section we discuss the average ε-complexity of linear problems. As in IBC, we make the following assumptions:

(a) simple operations on the elements of the space G, such as addition and multiplication by a scalar, cost unity;

(b) the cost of one functional evaluation (observation) is constant and equal to c.

For information N (in general adaptive) and an algorithm ϕ which uses N, let $\mathrm{cost}(N, f, z)$ be the information cost of obtaining the noisy information z of $N(f)$. In particular, $\mathrm{cost}(N, f, z) \geq cn(z)$, where $n(z)$ denotes, as before, the total number of noisy functional evaluations. Let $\mathrm{cost}(\phi, z)$ be the combinatory cost of computing $\phi(z)$. Then the average cost of obtaining approximation to Sf is given by

$$\mathrm{cost}_\delta(\phi, N) = \int_F \left(\int_Y (\mathrm{cost}(N, f, z) + \mathrm{cost}(\phi, z))\pi_\delta(dz \mid f, N)\right) \mu(df).$$

The average ε-complexity is defined as the minimal average cost of computing Sf with average error no greater than ε,

$$\mathrm{comp}_\delta(\varepsilon) = \inf\{\mathrm{cost}_\delta(\phi, N) \colon \phi, N \text{ such that } e_\delta(\phi, N) \leq \varepsilon\}.$$

Finally, the average ε-cardinality number $m^*_\delta(\varepsilon)$ is defined as

$$m^*_\delta(\varepsilon) = \inf\{\mathrm{card}_\delta(N) \colon N \text{ such that } r_\delta(N) \leq \varepsilon\}.$$

We are ready to show tight bounds on the average ε-complexity. We know from Theorem 3.1 that the optimal algorithm for nonadaptive information is linear. Hence, proceeding as in IBC for exact information we obtain

THEOREM 7.1. *The average ε-complexity of a linear problem with noisy information satisfies*

$$c \cdot m_\delta^*(\varepsilon) \le \mathrm{comp}_\delta(\varepsilon) \le (c + 2) \cdot m_\delta^*(\varepsilon).$$

The above theorem and Corollary 6.1 yield immediately

COROLLARY 7.1. *For any linear problem with noisy information we have*

$$\mathrm{comp}_\delta(\varepsilon) \ge c \cdot \delta \cdot (\lambda_1/\varepsilon^2 - 1).$$

For $\delta > 0$ and for any nonzero linear operator S, the corollary says that the average ε-complexity tends to infinity with $\varepsilon \to 0$ at least as fast as ε^{-2}.

EXAMPLE 7.1. We illustrate the average ε-complexity for $\lambda_j = (aj)^{-p}$, $j = 1, 2, \ldots,$ where $p > 1$. Using the formulas for $r_\delta^2(n)$ from Example 4.2 we obtain the estimate

$$\mathrm{comp}_\delta(\varepsilon) = (c + b) \cdot \vartheta(\delta, \varepsilon) \cdot (1 + o(1)) \qquad \text{as } \varepsilon \to 0,$$

where $0 \le b \le 2$ and the function $\vartheta(\delta, \varepsilon)$ is given below ($\delta \ge 0$):

$$p > 2: \quad \delta Z_{a,p}^2 \left(\frac{1}{\varepsilon}\right)^2 + \frac{a^{p/(1-p)}}{p-1}\left(\frac{1}{\varepsilon}\right)^{2/(p-1)}, \; Z_{a,p} = \sum_{j=1}^{\infty} (aj)^{-p/2}$$

$$p = 2: \quad \frac{\delta}{a^2}\left(\frac{1}{\varepsilon}\right)^2 \ln^2\left(\frac{1}{\varepsilon}\right) + \frac{1}{a}\left(\frac{1}{\varepsilon}\right)^2$$

$$p < 2, \; \delta \le (2-p)/p: \quad a^{p/(1-p)}\left(\frac{\delta}{1+\delta}\frac{1}{(1-p/2)^2} + \frac{1}{p-1}\right)^{1/(p-1)}\left(\frac{1}{\varepsilon}\right)^{2/(p-1)}$$

We now comment on this. For $p > 2$ the average ε-complexity for exact information, $\delta = 0$, becomes proportional to $\varepsilon^{-2/(p-1)}$. Thus for large p it tends to infinity rather slowly. The presence of noise, $\delta > 0$, changes the average ε-complexity to be proportional to $\delta\varepsilon^{-2}$. This is a significant change in the exponent since $2/(p - 1)$ is replaced by 2.

Let us turn to the case $p = 2$. For exact information the average ε-complexity is proportional to ε^{-2}, whereas for noisy information it is proportional to $\delta\varepsilon^{-2} \ln^2(1/\varepsilon)$. In this case the noise makes a slight change in the form of the average ε-complexity.

Finally consider $p < 2$. For exact information the average ε-complexity is proportional to $\varepsilon^{-2/(p-1)}$ which tends to infinity quite rapidly as ε approaches zero. In this case noisy information does not effect the average ε-complexity and the noise is almost irrelevant.

Remark 7.1. We know from Theorems 6.1 and 7.1 how to achieve the upper bound on the average ε-complexity. However, one may want to use only the μ-orthonormal functionals K_i^*, $i = 1, 2, \ldots$, in order to obtain ε-approximations to Sf.

Using Remark 4.1 we can show information N_ε consisting only of functionals K_i^*, $i = 1, 2, \ldots$, for which $r_\delta(N_\varepsilon) \le \varepsilon$, and

$$\mathrm{comp}_\delta(\varepsilon) \le \mathrm{cost}_\delta(\phi_{\delta,N_\varepsilon}, N_\varepsilon) \le 2\,\frac{c+2}{c}\,\mathrm{comp}_\delta(\varepsilon).$$

8. An Example

We illustrate how the results of this paper may be applied to the approximation problem in finite dimensional spaces.

Let $F = G = \mathbb{R}^d$, $d \ge 1$. Let μ be the Gaussian measure on $\mathbb{R}^d$ with mean zero and correlation operator λI, $\lambda > 0$; i.e.,

$$\mu(B) = (2\pi\lambda)^{-d/2} \int_B \exp\{-\|f\|^2/(2\lambda)\}\, df, \ \forall B \in \mathbb{B}(\mathbb{R}^d).$$

We assume that the variance σ_x^2 of the random variable corresponding to the functional $L_x = \langle \cdot, x\rangle$ is equal to $\gamma\|x\|_2^2$, $\gamma \ge 0$. The parameter δ is then equal to $\delta = \gamma/\lambda$.

Suppose we want to approximate the identity operator, $Sf = f$, using any one-to-one information operator N: $F \to \mathbb{R}^n$,

$$N(f) = [\langle f, x_1\rangle, \langle f, x_2\rangle, \ldots, \langle f, x_n\rangle], \qquad \|x_i\|_2 = 1, \qquad \forall i.$$

The optimal algorithm for N is of the form

$$\phi_{\delta,N}(\alpha) = \sum_{j=1}^{n} \beta_i x_i,$$

where β is the solution of the linear system

$$(\delta/\lambda \cdot I + M_N)\beta = \alpha, \qquad M_N = (\langle x_i, x_j\rangle)_{i,j=1}^n.$$

Its error is equal to

$$e_\delta^2(\phi_{\delta,N}, N) = \gamma \cdot \sum_{j=1}^{d} (\gamma/\lambda + \eta_j)^{-1},$$

where η_j, $j = 1, \ldots, d$ are the eigenvalues of the matrix M_N.

In problems like this the well known least-squares algorithm ϕ_N^{LS} is often used in practice. In the general case, this algorithm is defined as $\phi_N^{\mathrm{LS}}(\beta) = N^{-1}P_N\beta$, where P_N is the orthorgonal projection from $\mathbb{R}^n$ onto the subspace $N(F)$. The least-squares algorithm does not need the knowledge of the measure μ and parameter γ. That is why it seems interesting to compare the error of ϕ_N^{LS} with the error of the optimal algorithm. It turns out that the least-squares algorithm is close to optimal whenever the ratio γ/λ is small. Indeed, we have $\|f - \phi_N^{\mathrm{LS}}(\beta)\| = \|f - N^{-1}P_N\beta\| = \|f - N^{-1}P_N(Nf + (\beta - Nf))\| = \|N^{-1}P_N(\beta - Nf)\|$. Hence,

$$\begin{aligned}\int_{\mathbb{R}^n} \|f - \phi_N^{\mathrm{LS}}(\beta)\|^2 \pi_\delta(d\beta|f, N) &= \int_{\mathbb{R}^n} \|N^{-1}P_N\beta\|^2 \pi_\delta(d\beta|0, N)\\ &= \gamma \cdot \operatorname{trace}(N^{-1}P_N(N^{-1}P_N)^*)\\ &= \gamma \cdot \operatorname{trace}(N^*N)^{-1} = \gamma \cdot \sum_{j=1}^{d} \eta_j^{-1}.\end{aligned}$$

Finally, $e_\delta^2(\phi_N^{\mathrm{LS}}, N) = \gamma \cdot \sum_{j=1}^d \eta_j^{-1} = \lim_{\gamma/\lambda \to 0} r_\delta^2(N)$.

We now find the nth optimal average radius and ε-complexity for our problem. From Theorem 4.1 we get

$$r_\delta^2(n) = \gamma d^2/(n + \gamma d/\lambda), \qquad \forall n \geq d.$$

Hence, for $\varepsilon \leq \varepsilon_0 = \gamma d/(1 + \gamma/\lambda)$ we have

$$m_\delta(\varepsilon) \leq m_\delta^*(\varepsilon) \leq \lceil m_\delta(\varepsilon) \rceil,$$

where $m_\delta(\varepsilon) = \gamma d(d/\varepsilon^2 - 1/\lambda)$. Finally,

$$\operatorname{comp}_\delta(\varepsilon) = (c + b) \cdot \tau(\delta, \varepsilon) \cdot (1 + o(1)) \qquad \text{as } \varepsilon \to 0,$$

where $0 \leq b \leq 2$ and $\tau(\delta, \varepsilon) = \gamma d^2/\varepsilon^2$ for $\delta > 0$, $\tau(0, \varepsilon) = d$.

It is easy to check that for $n = ds$ the information

$$N_n = [\underbrace{N_0, N_0, \ldots, N_0}_{s}],$$

where $N_0 = [\langle \cdot, e_1\rangle, \langle \cdot, e_2\rangle, \ldots, \langle \cdot, e_d\rangle]$, is nth optimal. The optimal algorithm ϕ_n for N_n now takes the form

$$\phi_n(\beta) = \left((s + \delta/\lambda)^{-1} \cdot \sum_{j=1}^{s} \beta_{ij}\right)_{i=1}^{d}, \qquad \forall \beta \in \mathbb{R}^n,$$

where $\beta = N_n(f) + \alpha = [\beta_{11}, \beta_{12} \ldots, \beta_{1d}, \ldots, \beta_{s1}, \ldots, \beta_{sd}]$. On the other hand, the least-squares algorithm is

$$\phi_n^{\mathrm{LS}}(\beta) = \left(s^{-1} \cdot \sum_{j=1}^{s} \beta_{ij}\right)_{i=1}^{d}, \qquad \forall \beta \in \mathbb{R}^n.$$

Furthermore,

$$e_\delta^2(\phi_n^{\mathrm{LS}}, N_n) = \gamma \cdot d^2/n = r_\delta^2(n) \cdot (1 + \gamma d/(\lambda n)).$$

We conclude that the pair N_n, ϕ_n^{LS} with $n = \lceil \gamma d^2/\varepsilon^2 \rceil$ produces ε-approximation with the cost $(c + 2)\lceil \gamma d^2/\varepsilon^2 \rceil$. This cost is close to optimal.

Remark 8.1. We have shown that the least-squares algorithm is, in fact, not optimal. However, it is optimal in the mixed setting, where the error of an algorithm ϕ, which uses information N, is defined as

$$e_\gamma^{\mathrm{wor}}(\phi, N) = \sup_{f \in F} \sqrt{\int_{\mathbb{R}^n} \|f - \phi(Nf + \beta)\|^2 \chi_\gamma(d\beta)},$$

where χ_γ is Gaussian on $\mathbb{B}(\mathbb{R}^n)$ with mean zero and correlation operator γI. Indeed, since $e_\gamma^{\mathrm{wor}}(\phi, N) \geq e_\delta(\phi, N)$, $\forall \phi$, $\forall \lambda$, we get $e_\gamma^{\mathrm{wor}}(\phi_N^{\mathrm{LS}}, N) = e_\delta(\phi_N^{\mathrm{LS}}, N) = \lim_{\lambda \to \infty} r_\delta(N) \leq e_\gamma^{\mathrm{wor}}(\phi, N)$, $\forall \phi$. The error $e_\gamma^{\mathrm{wor}}(\phi, N)$ is thus minimal at $\phi = \phi_N^{\mathrm{LS}}$.

Appendix

In this appendix we present the conditional measures with respect to noisy information. We use notation from the paper.

Consider first nonadaptive information $N = [L_1, L_2, \ldots, L_n]$. Let $\bar{F} = F \times \mathbb{R}^n = \{(f, z): f \in F, z \in \mathbb{R}^n\}$. We may formally treat $\bar{F}$ as a normed space with the norm $\|(f, z)\|_*^2 = \|f\|^2 + \|z\|_2^2$. The information N generates the probability measure $\bar{\mu}$ on $\mathbb{B}(\bar{\mathbb{F}})$, defined by

$$\bar{\mu}(A \times B; N) = \int_A \pi_\delta(B|f, N)\mu(df), \ \forall A \in \mathbb{B}(F), \forall B \in \mathbb{B}(\mathbb{R}^n).$$

$$\text{Let } \mu_1(B; N) = \bar{\mu}(F \times B; N) = \int_F \pi_\delta(B|f, N)\mu(df), \ \forall B \in \mathbb{B}(\mathbb{R}^n).$$

Hence $\mu_1(B; N)$ is the probability that the computed information z belongs to B. Let $M_N = (\langle L_i, L_j\rangle_\mu)_{ij=1}^n$.

LEMMA A1. *For any nonadaptive noisy information $N = [L_1, \ldots, L_n]$, with $\|L_j\|_\mu = 1, \forall j = 1, 2, \ldots, n$, the measure $\mu_1(\circ; N)$ is Gaussian with mean zero and correlation operator C_N: $\mathbb{R}^n \to \mathbb{R}^n$, $C_N z = (\delta I + M_N)z$. Furthermore,*

$$\bar{\mu}(A \times B; N) = \int_B \mu_2(A|z, N)\mu_1(dz; N), \qquad \forall A \in \mathbb{B}(F), \forall B \in \mathbb{B}(\mathbb{R}^n),$$

where $\mu_2(\circ|z, N)$ is Gaussian on $\mathbb{B}(F)$ with mean $m(z) = \sum_{j=1}^n y_j(C_\mu L_j)$, where the vector $y = [y_1, \ldots, y_n]$ satisfies $(\delta I + M_N)y = z$, and correlation operator $C_{\mu,N}$: $F^* \to F$, given by

$$C_{\mu,N}L = C_\mu L - m(N(C_\mu L)), \qquad \forall L \in F^*.$$

Note that for $\delta = 0$ and linearly dependent functionals L_i the vector y is understood as in Theorem 3.1.

Proof. The measure $\pi_\delta(\circ|f, N)$ is Gaussian with mean $N(f)$ and correlation operator δI. Hence, for the characteristic functional ψ_N: $\mathbb{R}^n \to \mathbb{C}$ of the measure μ_1 we have ($i = \sqrt{-1}$)

$$\begin{aligned}\psi_N(w) &= \int_{\mathbb{R}^n} \exp\{i\langle w, z\rangle_2\}\mu_1(dz; N)\\ &= \int_F \left(\int_{\mathbb{R}^n} \exp\{i\langle w, z\rangle_2\}\pi_\delta(dz|f, N)\right) \mu(df)\\ &= \int_F \exp\left\{i\langle w, Nf\rangle_2 - \frac{1}{2}\langle \delta w, w\rangle_2\right\} \mu(df)\end{aligned}$$

$$= \exp\left\{-\frac{1}{2}\,\delta\|w\|_2^2\right\}\int_F \exp\{i\langle w, Nf\rangle_2\}\mu(df)$$

$$= \exp\left\{-\frac{1}{2}\,\delta\|w\|_2^2\right\}\int_F \exp\left\{i\sum_{j=1}^{n} w_j L_j(f)\right\}\mu(df)$$

$$= \exp\left\{-\frac{1}{2}\,\delta\|w\|_2^2\right\}\int_F \exp\left\{\frac{1}{2}\langle M_N w, w\rangle_2\right\}$$

$$= \exp\left\{-\frac{1}{2}\langle(\delta I + M_N)w, w\rangle_2\right\}.$$

The measure μ_1 is thus Gaussian with mean zero and correlation operator $\delta I + M_N$.

We now prove the second part of the lemma. As in IBC for exact information we can show that the measure $\mu_2(\circ|z, N)$ is well defined for any z. Furthermore, $\mu_2(B|z, N)$ is a μ_1-measurable function of z for any Borel set B. To complete the proof it is therefore enough to show that the characteristic functional $\bar{\psi}$ of $\bar{\mu}$ is equal to the characteristic functional $\bar{\psi}'$ of the measure μ' defined as

$$\mu'(A \times B) = \int_B \mu_2(A|z, N)\mu_1(dz; N), \qquad \forall A \in \mathbb{B}(F), \forall B \in \mathbb{B}(\mathbb{R}^n).$$

To this end, let $\bar{L} \in \bar{F}^*$. Then there exist $L \in F^*$ and $w \in \mathbb{R}^n$ such that $\bar{L}(f, z) = Lf + \langle w, z\rangle_2$, $\forall f \in F$, $\forall z \in \mathbb{R}^n$. We have

$$\bar{\psi}'(\bar{L}) = \int_{\mathbb{R}^n}\left(\int_F \exp\{i(Lf + \langle w, z\rangle_2)\}\,\mu_2(df|z, N)\right)\mu_1(dz; N)$$

$$= \int_{\mathbb{R}^n}\exp\{i\langle w, z\rangle_2\}\left(\int_F \exp\{iL(f)\}\mu_2(df|z; N)\right)\mu_1(dz; N)$$

$$= \int_{\mathbb{R}^n}\exp\left\{i\left(\langle w, z\rangle_2 + L(m(z)) - \frac{1}{2}(L(C_\mu L) - L(m(NC_\mu L)))\right)\right\}$$

$$\times\ \mu_1(dz; N).$$

Observe that $L(m(z)) = L(\sum_{j=1}^n y_j(C_\mu L_j)) = \sum_{j=1}^n y_j L_j(C_\mu L) = \langle(\delta I + M_N)^{-1}z, NC_\mu L\rangle_2 = \langle z, (\delta I + M_N)^{-1}NC_\mu L\rangle_2$. Similarly, $L(m(NC_\mu L)) = \langle NC_\mu L, (\delta I + M_N)^{-1}NC_\mu L\rangle_2$. Hence,

$$\begin{aligned}\psi'(L') &= \exp\left\{-\frac{1}{2}\,(L(C_\mu L) - \langle NC_\mu L, (\delta I + M_N)^{-1}NC_\mu L\rangle_2)\right\}\\ &\quad \int_{\mathbb{R}^n} \exp\{i\langle z, w + (\delta I + M_N)^{-1}NC_\mu L\rangle_2\}\mu_1(dz; N)\\ &= \exp\left\{-\frac{1}{2}\,(L(C_\mu L) - \langle NC_\mu L, (\delta I + M_N)^{-1}NC_\mu L\rangle_2)\right\}\\ &\quad \exp\left\{-\frac{1}{2}\,(\langle(\delta I + M_N)w, w\rangle_2\right.\\ &\quad \left. + \langle NC_\mu L, (\delta I + M_N)NC_\mu L\rangle_2 + 2\langle w, NC_\mu L\rangle_2)\right\}\\ &= \exp\left\{-\frac{1}{2}\,(L(C_\mu L) + 2\langle w, NC_\mu L\rangle_2 + \langle(\delta I + M_N)w, w\rangle_2)\right\}.\end{aligned}$$

On the other hand, for the characteristic functional $\bar{\psi}$ of the measure $\bar{\mu}$ we have

$$\begin{aligned}\bar{\psi}(\bar{L}) &= \int_F \left(\exp\{i(Lf + \langle w, z\rangle_2)\}\pi_\delta(dz \mid f, N)\right)\mu(df)\\ &= \int_F \exp\{iLf\}\left(\int_{\mathbb{R}^n} \exp\{i\langle w, z\rangle_2\}\pi_\delta(dz \mid f, N)\right)\mu(df)\\ &= \exp\left\{-\frac{1}{2}\,\delta\|w\|_2^2\right\}\int_F \exp\{i(Lf + \langle Nf, w\rangle_2)\}\mu(df)\\ &= \exp\left\{-\frac{1}{2}\,(L(C_\mu L) + 2\langle w, NC_\mu L\rangle_2 + \langle(\delta I + M_N)w, w\rangle_2)\right\}\\ &= \bar{\psi}'(\bar{L}).\end{aligned}$$

This completes the proof. ■

We now turn to adaptive information N^{a},

$$N^{\mathrm{a}}(f) = [L_1(f), L_2(f; z_1), \ldots, L_{n(z)}(f; z_1, \ldots, z_{n(z)-1})].$$

In this case the measure $\bar{\mu}(\circ; N^{\mathrm{a}})$ is defined on the space $\bar{F} = F \times Y$, where $Y = \bigcup_{m=1}^{\infty} B_m$ and B_m are as in Section 4. Furthermore,

$$\bar{\mu}(A \times B) = \int_A \pi_\delta(B \mid f, N^{\mathrm{a}})\mu(df)$$

$$= \sum_{m=1}^{\infty} \int_A \pi_\delta(B \cap B_m | f, N^a)\mu(df),$$

$$\forall A \in \mathbb{B}(F), \forall B \in \mathbb{B}(Y)$$

The measure $\mu_1(\circ; N^a)$ is now given by

$$\mu_1(B; N^a) = \bar{\mu}(\bar{F} \times B) = \sum_{m=1}^{\infty} \int_F \pi_\delta(B \cap B_m | f, N^a)\mu(df).$$

Note that, unlike in the nonadaptive case, the measure μ_1 is, in general, non-Gaussian.

LEMMA A2. *For any adaptive noisy information*

$$N^a(f) = [L_1(f), L_2(f; z_1), \ldots, L_{n(z)}(f; z_1, \ldots, z_{n(z)-1})],$$

with $\|L_j(\circ; z_1, \ldots, z_{j-1})\|_\mu = 1$, $\forall j = 1, 2, \ldots, \forall z \in Y$, *we have*

$$\bar{\mu}(A \times B; N) = \int_B \mu_2(A | z, N_z^{\text{non}})\mu_1(dz; N^a)$$

$$= \sum_{m=1}^{\infty} \int_{B \cap B_m} \mu_2(A | z, N_z^{\text{non}})\, \mu_1(dz; N^a),$$

where $N_z^{\text{non}} = [L_1(\circ), L_2(\circ; z_1), \ldots, L_{n(z)}(\circ; z_1, \ldots, z_{n(z)-1})]$.

Proof. We first prove the lemma for any information N^a with fixed cardinality, i.e., $n(z) \equiv n$. This will be done by induction on n.

For $n = 1$ the proof is obvious. Suppose that $n > 1$. Let N_{n-1}^a be the adaptive information consisting of the first $(n - 1)$ functionals from N^a. Then for $A \in \mathbb{B}(F)$, $B \in \mathbb{B}(\mathbb{R}^{n-1})$, $C \in \mathbb{B}(\mathbb{R})$ we have

$$\bar{\mu}(A \times B \times C) = \int_A \pi_\delta(B \times C | f, N^a)\mu(df)$$

$$= \int_A \left(\int_B \pi_\delta(C | f, L_{n,t})\pi_\delta(dt | f, N_{n-1}^a)\right) \mu(df)$$

$$= \int_B \left(\int_A \pi_\delta(C | f, L_{n,t})\mu_2(df | t, N_t^{\text{non}})\right) \mu(dt; N_{n-1}^a),$$

where

$$t = [t_1, t_2, \ldots, t_{n-1}] \qquad \text{and} \qquad L_{n,t} = L_n(\circ; t).$$

Calculations similar to those from the proof of Lemma A1 yield that

$$\int_A \pi_\delta(C|f, L_{n,t})\, \mu_2(df|t, N_t^{\mathrm{non}}) = \int_C \mu_2(A|[t, t_n], N_{[t,t_n]}^{\mathrm{non}})\omega(dt_n; L_{n,t}),$$

where ω is one dimensional Gaussian measure defined as

$$\omega(D; L_{n,t}) = \int_{\mathbb{R}} \pi_\delta(D|f, L_{n,t})\mu_2(df|t, N_t^{\mathrm{non}}), \qquad \forall D \in \mathbb{B}(\mathbb{R}).$$

Thus, to complete the induction step it is enough to observe that

$$\begin{aligned}
&\int_B \left(\int_C \omega(dt_n; L_{n,t})\right) \mu_1(dt; N_{n-1}^{\mathrm{a}}) \\
&= \int_B \omega(C; L_{n,t})\mu_1(dt; N_{n-1}^{\mathrm{a}}) \\
&= \int_B \left(\int_F \pi_\delta(C|f, L_{n,t})\mu_2(df|t, N_t^{\mathrm{non}})\right) \mu_1(dt; N_{n-1}^{\mathrm{a}}) \\
&= \int_F \pi_\delta(B \times C|f, N^{\mathrm{a}})\mu(df) = \mu_1(B \times C; N^{\mathrm{a}}).
\end{aligned}$$

Now, let N^{a} be arbitrary adaptive information. Then, for $A \in \mathbb{B}(F)$, $B \in \mathbb{B}(Y)$ we have

$$\begin{aligned}
\tilde{\mu}(A \times B) &= \sum_{m=1}^{\infty} \int_A \pi_\delta(B \cap B_m|f, N^{\mathrm{a}})\mu(df) \\
&= \sum_{m=1}^{\infty} \int_B \mu_2(A|z, N_z^{\mathrm{non}})\mu_1(dz; N^{\mathrm{a}}),
\end{aligned}$$

as claimed. ■

ACKNOWLEDGMENTS

I thank H. Woźniakowski for his invaluable help during the preparation of this paper.

REFERENCES

CHALONER, K. (1984), Optimal Bayesian experimental designs for linear models, *Ann. Statist.* **12,** 283–300.

CRAVEN, P., AND WAHBA, G. (1979), Smoothing noisy data with spline functions: Estimating the current degree of smoothing by the method of generalized cross-validation, *Numer. Math.* **31,** 377–403.

FEDOROV, V. V. (1972), "Theory of Optimal Experiments," Academic Press, New York.

KACEWICZ, B. Z., MILANESE, M., TEMPO, R., AND VICINO, A. (1986), Optimality of central and projection algorithms for bounded uncertainty, *Systems Control Lett.* **8,** 161–171.

KADANE, J. B., AND WASILKOWSKI, G. W. (1985), "Average Case ε-complexity in Computer Science—A Bayesian View," Elsevier/North-Holland, Amsterdam.

KADANE, J. B., WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1989), On adaption with noisy information, *J. Complexity,* in press.

KIEFER, J. (1959), Optimal experimental designs (with discussion), *J. Roy. Statist. Soc. Ser. B,* **21,** 272–319.

KIEFER, J., AND WOLFOWITZ, J. (1959), Optimal designs in regression problems, *Ann. Math. Statist.* **30,** 271–294.

KIMELDORF, G. S., AND WAHBA, G. (1970), A correspondence between Bayesian estimation of stochastic processes and smoothing by splines, *Ann. Math. Statist.* **41,** 495–502.

KUO, H. H. (1975), "Gaussian Measures in Banach Spaces," Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin/New York.

LEE, D., PAVLIDIS, T., AND WASILKOWSKI, G. W. (1987), A note on the trade-off between sampling and quantization in signal processing, *J. Complexity* **3,** 359–371.

MARCHUK, A. G., AND OSIPENKO, K. YU. (1975), Best approximation of functions specified with an error at a finite number of points, *Math. Notes* **17,** 207–212.

MELKMAN, A. A., AND MICCHELLI, C. A. (1979), Optimal estimation of linear operators in Hilbert spaces from inaccurate data, *SIAM J. Numer. Anal.* **16,** 87–105.

MICCHELLI, C. A., AND RIVLIN, T. J. (1977), A survey of optimal recovery, *in* "Estimation in Approximation Theory," pp. 1–54, Plenum, New York.

MILANESE, M., AND TEMPO, R. (1984), Optimal algorithm theory for robust estimation and prediction, *IEEE Trans. Automat. Control* **AC-30,** 230–238.

O'HAGAN, A. (1978), Curve fitting and optimal design for prediction (with discussion), *J. Roy. Statist. Soc. Ser. B* **40,** 1–41.

RAGOZIN, D. L. (1983), Error bounds for derivative estimates based on spline smoothing of exact or noisy data, *J. Approx. Theory* **37,** 335–355.

REINSCH, C. H. (1971), Smoothing by spline functions, II. *Numer. Math.* **16,** 451–454.

SACKS, J., AND YLVISAKER, D. (1966), Designs for regression problems with correlated errors, *Ann. Math. Statist.* **37,** 66–89.

SPECKMAN, P. (1985), Spline smoothing and optimal rates of convergence, in nonparametric regression models, *Ann. Statist.* **13,** 970–983.

TRAUB, J. F., WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1988), "Information-Based Complexity," Academic Press, New York.

VAKHANIA, N. N. (1981), "Probability Distributions on Linear Spaces," North-Holland, New York.

VICINO, A., TEMPO, R., GENESIO, R., AND MILANESE, M. (1987), Optimal error and GMDH predictors: A comparison with some statistical techniques, *Internat. J. Forecasting,* to appear.

WAHBA, G. (1971), On the regression design problems of Sacks and Ylvisaker, *Ann. Math. Statist.* **42,** 1035–1053.

WAHBA, G. (1975), Smoothing noisy data by spline functions, *Numer. Math.* **24,** 383–393.

WAHBA, G. (1985), A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, *Ann. Statist.* **13,** 1378–1402.

WAHBA, G. (1984), Cross-validated spline methods for the estimation of multivariate functions from data on functionals, *in* "Stochastics: An appraisal, Proceedings 50th Anniversary Conference, Iowa State Statistics Laboratory," pp. 205–235, Iowa State Univ. Press, Ames.

WAHBA, G., AND WANG, Y. (1987), "When Is the Optimal Regularization Parameter Insensitive to the Choice of the Loss Function?," Report, University of Wisconsin, Department of Statistics.

WASILKOWSKI, G. W. (1986), Information of varying cardinality, *J. Complexity* **2,** 204–225.

WASILKOWSKI, G. W. (1987), "A Clock Syncrhonization Problem with Random Delays," Report, Columbia University Computer Science Department; *J. Complexity,* to appear.